The Knowledge Burden Hypothesis: Evidence From the Collapse of the Soviet Union

Ajay Agrawal, Avi Goldfarb and Florenta Teodoridis*

June 2012

Preliminary: Please do not cite or circulate

Abstract

The knowledge burden hypothesis, that reaching the knowledge frontier is progressively costly and results in increasing specialization and thus increasing collaboration, has important implications for long-run growth and rising inequality. However, the hypothesis is difficult to bring to data due to identification challenges. We exploit the collapse of the USSR as an instrument since Soviet research was kept hidden from the outside world and then suddenly released, creating a shock to the frontier, due to socio-political events exogenous to the research community. Furthermore, the degree of the shock varied across subfields since Soviet advances were greater in some subfields than others. Using 40 years of publication data in theoretical mathematics covering the period 1970-2010, we employ a difference-in-differences type approach to compare the propensity of mathematicians working outside the USSR to collaborate in "Soviet-rich" versus "Soviet-poor" subfields before and after the knowledge frontier shock. Consistent with the theory, we find that team size - the number of coauthors on a paper - increases disproportionately more in Soviet-rich subfields after 1990.

JEL: O33, O40

* Agrawal: University of Toronto and NBER, 105 St. George St. Toronto, ON M5S 3E6, ajay.agrawal@rotman.utoronto.ca. Goldfarb: University of Toronto and NBER, 105 St. George St. Toronto, ON M5S 3E6, agoldfarb@rotman.utoronto.ca. Teodoridis: University of Toronto, 105 St. George St. Toronto, ON M5S 3E6, Florenta.Teodoridis09@rotman.utoronto.ca. This research was funded by the Centre for Innovation and Entrepreneurship at the Rotman School of Management, the Martin Prosperity Institute, and the Social Sciences and Humanities Research Council of Canada. We thank participants to the "The Organization, Economics and Policy of Scientific Research" workshop for valuable feedback. Errors remain our own. © 2012 by Ajay Agrawal, Avi Goldfarb and Florenta Teodoridis. All rights reserved.

1. Introduction

Benjamin Jones (2009) proposes a provocative theory that successive generations of innovators face an increasing education burden due to the constantly advancing knowledge frontier, which requires innovators to specialize more and thus necessitates working more collaboratively over time. This alters the organization of innovative activity towards more teamwork. The theory yields a meaningful implication: the nature of innovation is changing in a manner that has negative implications for long-run growth (Jones, 2009) and rising inequality (Jones, 2011).

Jones provides descriptive statistics consistent with this theory. For example, he shows that over time: 1) the number of co-authors on academic publications increases, 2) Nobel laureates are older when they perform their great achievement, 3) the number of co-inventors per patent increases, 4) the age at first innovation increases, and 5) the probability of switching fields decreases. However, these statistics are also consistent with other explanations, such as institutional changes (e.g., college requirements, journal publication processes; Stephan, 2012), reduced collaboration costs (Agrawal and Goldfarb, 2008), and increasing capital costs (Mowery et al, 2004).

Identification is difficult since many unobservables may be (and likely are) correlated with both collaborative behavior and time. In order to provide more compelling (albeit not definitive) evidence that an increasing knowledge burden leads to a growing propensity to collaborate, we need an instrument that is correlated with the knowledge frontier but not correlated with academic collaboration except indirectly through its effect on the frontier.

The collapse of the Soviet Union in 1989 provides such an instrument. Although the USSR was a world leader in certain areas of science, including various subfields of mathematics, Communist government officials forced their researchers to work in isolation from the rest of the world. For example, with few exceptions scholars were strictly prohibited from traveling, publishing outside of the Soviet Union, and accessing foreign publications without case-by-case government approval. Thus, when the Iron Curtain fell and Soviet science became widely available, the knowledge frontier outside the USSR experienced a shock.

Furthermore, the degree of the knowledge shock across subfields of mathematics varied since Soviet researchers had made significant advances relative to the rest of the world in some subfields but not others. We focus on theoretical mathematics. Borjas and Doran (2012) show that the Soviet mathematics community was very advanced relative to the West in some subfields of theoretical mathematics, such as "partial differential equations" and "operator theory," and much less so in others, such as "abstract harmonic analysis" and "sequences, series, summability."

We exploit this variation in knowledge shock, which is exogenous to the mathematics community, across subfields using a difference-in-differences type of analysis. Specifically, we compare the propensity of mathematicians working outside the USSR to collaborate in "Soviet-rich" versus "Soviet-poor" subfields before and after the shock to the knowledge frontier. We do this using 40 years of publication data in theoretical mathematics covering the period 1970-2010, 20 years before and after the collapse of the Soviet Union.

We categorize papers using the internationally recognized Mathematics Subject Classification codes developed and assigned by the Mathematical Reviews division of the

American Mathematical Society. We follow the Soviet-rich versus Soviet–poor subfield classification developed by Borjas and Doran (2012), which is based on the fraction of publications produced by Soviet researchers during the period 1970-1989. We then focus our attention on mathematicians working outside the USSR and drop observations that involve collaboration with Soviet researchers.

We find that team size - the number of coauthors on a paper - increases after the fall of the Iron Curtain, in both Soviet-rich and Soviet-poor subfields. However, consistent with the theory, team size grows disproportionately more in Soviet-rich subfields after the shock to the frontier. These results are robust to various cut-off points for Soviet-rich and Soviet-poor subfields of theoretical mathematics.

2. Historical Context

The former Soviet Union was a world-renowned center of scientific research with mathematics holding a prominent position. MIT professor Loren Graham, who dedicated his career to the study of history of science and technology in Russia, notes in his 2008 book: "Mathematics was probably the Soviet Union's strongest field, and the Moscow School of Mathematics its greatest achievement." Scholarly research in mathematics was highly regarded in the Soviet Union. It conferred status and prestige, attracted great minds, and offered financial conditions superior to many other Soviet Union occupations.

Soviet mathematics was strong across the entire spectrum of theoretical and applied mathematics. However, pure theory probably contained the greatest advancements (Graham, 1993). The reason rests mainly with the former USSR government's policies, which were heavy

on secrecy concerns and focused on maintaining strict control over technological development. Thus, it was politically easier for Soviet mathematicians to persistently build on their progress in pure theory than in areas where technology implementation was more immediate. Many great advancements in applied mathematics ended up being stalled inside the USSR or exploited further predominantly outside the USSR[1] (Graham, 1993).

In fact, studying the policies of the former USSR government alongside the history of mathematics generally are the main ways to understand the evolution of Soviet mathematics. Path dependency played a huge role in shaping Soviet research trajectories (Borjas and Doran, 2012). For instance, the success of Moscow mathematics can be traced back to Ergorov and his student N. N. Luzin (Tikhomirov, 2007). Luzin, whose famous work was mainly focused on the theory of functions, mentored subsequent generations of eminent Soviet mathematicians like A. N. Kolmogorov, who "was probably the best-known Soviet researcher on probability" (Graham, 1993). On the other hand, little outstanding mentorship was available to practitioners of other subfields of theoretical mathematics like algebraic geometry (Borjas and Doran, 2012). In general, one can argue that historical dependency led to heterogeneity in the academic strength and excellence in the research trajectories of mathematics (Borjas and Doran, 2012). Borjas and Doran (2012) rank Soviet contributions in mathematics between 1970 and 1989 and observe strong activity in various subfields of mathematical analysis and relatively weak activity in certain algebra subfields.

Adding to these historical dependency factors is the USSR political environment. Government policies caged Soviet knowledge in mathematics, implicitly generating segregation

[1] Exceptions are directly linked to government interests, such as the space program (Graham, 1993).

of research paths. Although the broad concept of science was in line with the Soviet ideology (Graham and Dezhina, 2008), free communication with the outside world, particularly Western countries, was not. The Communist government kept very strict control on international travel, regardless of reasons. Academics were asked to go through a very stringent[2] and lengthy process in order to have a chance at traveling to conferences outside the USSR. Even so, most approvals were granted only for travel in Eastern Europe (Ganguli, 2011).

Obtaining academic publications from outside the USSR was equally painful. Scientists had to place requests with the government for select publications. The Communist party would approve or deny the request based on political criteria that were frequently unrelated to academic considerations, for instance suspicions of negative connections with certain internal political beliefs. Even when an external publication was approved, the government handled the translation process (Polyak, 2001). Attempts to publish or communicate with academics outside the Soviet Union were also subject to similarly harsh scrutiny (Polyak, 2001; Borjas and Doran, 2012). Although these restrictions applied to all countries outside the USSR, the Soviet Union was particularly adamant in severing communication with Western countries. This fact is salient in understanding the caging effect on Soviet mathematics when acknowledging that all other prestigious mathematical research was located in the Western world. Specifically, the US accounted for 51% of total research in mathematics in the 1980s, followed by France with 10% and West Germany with 6.5% (Dubois et al, 2011).

Not only did Soviet knowledge in mathematics take shape with very limited influence from and awareness of related undertakings outside the USSR, but it also remained in relative

[2] The government required many personal details. Those with "tainted" backgrounds were blacklisted for travel (Polyak, 2001).

secrecy until the collapse of the Soviet Union. First, the USSR government kept much of Soviet science secret (Graham and Dezhina, 2008). Second, what escaped the secrecy filter was subject to the natural barrier imposed by the Russian language. Graham and Dezhina (2008) note "the Russian language was known by few researchers outside the Soviet Union, and consequently the achievements of Soviet researchers were more frequently overlooked than those presented in more accessible languages."

Furthermore, mathematics is an area where scholars rely heavily on face to face interaction. Walsh and Bayma (1996) report that mathematicians are the scholars who need to travel the most out of all science disciplines. Virtual communication is employed only after mathematicians meet in person a few times and have a project well under way. This peculiarity of mathematics research exacerbates the magnitude of Soviet government restrictions implications for Soviet mathematicians. Travel restrictions for Soviet mathematicians had a huge impact on their propensity to collaborate with non-Soviet scholars, especially Western scholars. Thus Soviet mathematics did indeed develop in isolation.

What is also interesting is that the imposed seclusion did not negatively affect the outstanding quality of Soviet mathematics. Citing Lipman Bers, former president of the American Mathematical Society, Graham and Dezhina (2008) emphasize the strength of Soviet mathematics, which rested on deep-rooted talent, unmoved in its excellence by political oppression.

In sum, we argue that the fall of the Iron Curtain provides an excellent natural experiment affecting the knowledge frontier in certain subfields of theoretical mathematics. This unanticipated (among academics) historical event set free a large pool of accumulated

knowledge to contribute to global developments in mathematics. Borjan and Doran (2012) cite comprehensive evidence attesting to the fact that the historical event took the global mathematics community by surprise; even in the late 1980s, both the Western mathematical community and USSR scholars were quite certain that Soviet mathematics would remain secluded for the foreseeable future.

3. Estimation Strategy

We base our main estimation strategy on the difference-in-differences approach. The difference-in-differences identification allows for comparison of co-authorship rates of subfields affected by Soviet knowledge (treated) matched against Soviet-poor subfields (controls) before and after the fall of the Iron Curtain. This approach is superior to a simple cross-sectional estimation, which would fail to account for the unobserved heterogeneity of mathematical subfields. We craft the difference-in-differences estimation such that it distinguishes between the rise in team size directly attributable to the shock (the fall of the Iron Curtain) and the rise in team size determined by unobserved differences between treated and control subfields (selection).

Formally, we examine how the authorship team size of Soviet-rich subfields (treated) differs relative to the team size of matched Soviet-poor subfields (controls). We estimate log linear regressions at the publication-year level:

$$TeamSize_{it} = \beta(SovietRich_i \times AfterIronCurtain_t) + Subfield_i + \gamma_t + \varepsilon_{it}$$

$TeamSize_{it}$ is the logged count of authors for each academic paper *i* published in year t. $SovietRich_i$ is an indicator variable equal to 1 if the publication *i* belongs to the treated group and 0 otherwise. $AfterIronCurtain_t$ is an indicator variable equal to 1 if the publication *i* is released in a year after the fall of the Iron Curtain and 0 otherwise. This applies to publications of both treated and control subfields. We calculate the key explanatory variable as the interaction between $SovietRich_i$ and $AfterIronCurtain_t$: equal to 1 for those publications *i* that belong to the treated group of subfields and that were released after the fall of the Iron Curtain and equal to 0 for all others. We include subfield and time fixed effects, hence the individual terms $SovietRich_i$ and $AfterIronCurtain_t$ are not part of the estimating equation. For our data to reject the knowledge burden hypothesis, β must not be positive and significant. In other words, an estimated value of β that is positive implies that the average team size in Soviet-rich subfields did indeed increase disproportionately, relative to Soviet-poor subfields, after the knowledge shock caused by the collapse of the Soviet Union, consistent with the theory.

4. Data

4.1. Data scope

We collect the data from the American Mathematical Society (AMS). The Mathematical Reviews (MR) division of AMS maintains a comprehensive bibliographic database of worldwide academic publications in mathematics. The MR team is painstakingly engaged in classifying each paper uploaded to their database using the Mathematics Subject Classification (MSC) codes. The MSC schema is internationally accepted and used to identify global advancements in

subfields of mathematics. It facilitates targeted searches on research subjects across all subfields of mathematics. The MR team assigns precisely one primary MSC code to each academic publication uploaded to the MR database. We use this classification to isolate and compare the publication activity for Soviet-rich subfields with Soviet-poor subfields.

The MR database keeps track of all mathematics-related journal publications covering the three main categories of mathematics: mathematical foundations (including history and biography), pure or theoretical mathematics, and applied mathematics.[3] Our focus is on theoretical mathematics, which includes analysis, algebra, and geometry (Figure 1).

4.2. Dataset construction

4.2.1. Identifying Soviet-rich subfields of theoretical mathematics

A total of 40 active primary MSC codes (14 algebra, 19 analysis, 7 geometry) comprise the theoretical mathematics group. In order to assess the impact of former USSR knowledge on authorship team size, we need to identify those subfields of Soviet specialty before the fall of the Iron Curtain. In other words, we need to identify those subfields where the knowledge frontier was most impacted. This is a multi-step process.

First, we set the cut-off point for secluded Soviet knowledge build-up to 1990. The high-intensity political and social turmoil preceding and following the fall of the Iron Curtain spanned a period of roughly three years, between 1989 and 1991. We follow Borjas and Doran's (2012) interpretation of historical events, which isolates 1990 as the year when academic seclusion was significantly lessened. In order to ensure data sample variation, we aim to cover the period

[3] There are various ways of grouping research areas in mathematics. The one mentioned here follows the MSC code sequencing.

between 1970 and 2010, twenty years both before and after the fall of the Iron Curtain.

Second, we start constructing the pre-1990 ranking of subfields of Soviet specialty. We identify 34 subfields that span our period of interest. We eliminate one subfield for which we have incomplete data, for a total of 33 subfields. Next, we turn to Borjas and Doran's (2012) classification of Soviet contributions to mathematics. The authors construct the ranking using MR data counts of Soviet publications weighted against US publications between 1970 and 1989. It is important to note that pre-1990 Soviet publications were added to the MR database after the fall of the Iron Curtain. Prior to the fall these publications were kept secret within the Soviet Union as argued above. Hence, when constructing the ranking, Borjas and Doran (2012) exploit the fact that the American Mathematical Society has collected and categorized pre-1990 publications once access restrictions to these publications were lifted by the fall of the Iron Curtain.

Borjas and Doran (2012) define Soviet publications as journal publications with at least one Soviet author. Borjas and Doran (2012) accurately identify Soviet authors using the unique identifier the MR team assigns to each scholar who publishes in mathematics journals. We extract and normalize the ranking for the 33 subfields of interest.

However, Borjas and Doran (2012) rank Soviet contributions relative to American contributions, while our goal is to rank Soviet contributions relative to all other non-Soviet contributions across the 33 subfields of theoretical mathematics. We argue that the ranking holds well in a global context since the US had the largest share of contributions in mathematics, accounting for as noted above 51% of total publications between 1970 and 1989 (Dubois et al, 2011). Nevertheless, we conduct a robustness check of this claim by constructing

our own classification using an algorithm that identifies Soviet mathematicians based on name[4]. We recognize that this approach is noisier than the one employed by Borjas and Doran (2012) since we identify Soviet mathematicians based on last name, while Borjas and Doran (2012) rely on accurate institutional affiliation. However, we cannot mimic their methodology on global data since we do not have the same level of MR data access. We check if our name-based ranking classifies the 33 subfields within the same top/middle/bottom groups as Borjas and Doran (2012). We obtain a 100% mapping when considering each of the top, middle, and bottom groups spanning about a third of the ranking. We argue that this mapping is satisfactory for our analysis.

In what follows, we rely on the ranking of Soviet contribution to the 33 subfields of theoretical mathematics as constructed by Borjas and Doran (2012) (Table 1).

4.2.2. Avoiding confounding effects

We eliminate all Soviet journal publications defined as publications with at least one Soviet author. We do so to account for potential confounding effects. After 1990, not only was Soviet knowledge set free to contribute to global advancements in mathematics, but also collaboration restrictions were eliminated for Soviet mathematicians. By excluding all publications with at least one Soviet author, we account for the possibility of increased co-authorship rates due to the abolition of Soviet collaboration constraints.

[4] We identify Soviet last names based on conversations with experts and documented rules regarding Soviet surname endings. We then test and calibrate our algorithm by manually looking up and verifying if academics identified as having Soviet last names are indeed Soviets.

We identify Soviet authors using our name-based algorithm. We tag and eliminate all journal publications between 1970 and 2010 that contain at least one author identified as having a Soviet last name.

4.3. Descriptive statistics

We now have our 33 subfields of theoretical mathematics ranked based on the strength of Soviet contributions before the collapse of the Soviet Union which add up to a total of 563,462 publications spanning the 40 years period of interest.

Our empirical analysis rests on contribution variation across the 33 subfields since Soviet scholars published in all 33 subfields. Hence, by construction, the ranking does not provide a clear boundary between Soviet-rich and Soviet-poor subfields within the group of 33 subfields.

Thus we next turn to selecting such subfields. Generally, we consider high-ranking subfields to represent Soviet-rich subfields and low-ranking ones to represent Soviet-poor subfields. In our main specification we define Soviet-rich subfields as the top five ranking subfields of theoretical mathematics. Soviet-poor subfields are the bottom five ranking subfields of theoretical mathematics.

The results remain robust to alternative definitions of Soviet-rich. Specifically, we have also ran estimates with Soviet-rich defined as top three ranking subfields of theoretical mathematics and Soviet-poor defined as the bottom three ranking subfields of theoretical mathematics. In another set of robustness checks we define Soviet-rich as top quartile ranking subfields of theoretical mathematics as well as top half ranking subfields of theoretical

mathematics. Our results remain robust to all these alternative definitions of Soviet-rich. Furthermore, our results remain robust when considering a continuous measure of Soviet-rich.

For our main specification, where Soviet-rich is defined as top five ranking subfields of theoretical mathematics and Soviet-poor is defined as bottom five ranking subfields of theoretical mathematics, we calculate the mean team size and standard deviation before and after the collapse of the Soviet Union. The mean authorship team size for the treated subfields (Soviet-rich) is 1.320 for the 20-year period before 1990 and increases to 1.745 for the 20-year period after, with the standard deviation changing from 0.568 to 0.811. The control subfields (Soviet-poor) have a mean authorship team size of 1.282 with a standard deviation of 0.536 before 1990 and 1.621 with a standard deviation of 0.781 after 1990. There are a total of 263,533 publications in this sample spanning the 40 years period of interest.

We observe a disproportionate increase in team size for Soviet-rich subfields of theoretical mathematics after the collapse of the Soviet Union (Figure 2). Before 1990 the mean team size for Soviet-rich subfields was just 3% higher than the mean team size for Soviet-poor subfields of theoretical mathematics. We find that after the collapse of the Soviet Union this difference increases at 8%. This finding is in line with our prediction, providing support for the knowledge burden hypothesis. This is a significant increase, as theoretical mathematics generally exhibits the lowest average team size among the various areas of scientific research (Stephan, 2012).

However there is the concern that there are systematic differences between Soviet-rich and Soviet-poor subfields of theoretical mathematics. Hence in what follows we turn to the difference-in-differences estimation.

5. Estimating the knowledge burden hypothesis

We apply the difference-in-differences estimation to examine if the sudden knowledge boost in Soviet-rich theoretical mathematics subfields led to a higher increase in authorship team size than for Soviet-poor subfields. We present evidence that the influx of Soviet knowledge influenced a heterogeneous rise in co-authorship among mathematics scholars publishing in theoretical mathematics, as predicted by the theory.

Table 2 shows the difference-in-differences estimation results. In the first column we report difference-in-differences regression results for our main specification. In the following columns we include difference-in-differences regression results for alternative definitions of Soviet-rich. Specifically, in the second column we report estimate results for Soviet-rich defined as top three ranking subfields of theoretical mathematics and Soviet-poor defined as bottom three ranking subfields of theoretical mathematics. In next three columns we report estimates using the full dataset. First we define Soviet-rich as those subfields of theoretical mathematics which rank in the top quartile, while all other subfields comprise the Soviet-poor set. Next we extend the Soviet-rich range to include those subfields of theoretical mathematics which rank in the top 50 percentile. Last, we use a continuous measure of Soviet-rich based on reverse ranking position. Specifically $SovietRich_i$ is calculated as the negative value of the ranking position, such that top ranking subfields are assigned the higher numerical values. For example the top one ranking subfield is assigned a value of minus one and second ranking subfield is

assigned a value of minus two. This approach ensures that the top one ranking subfield has indeed a higher numerical value than the second ranking subfield of theoretical mathematics.

As Table 2 shows, the estimated value of β is positive and significant for all definitions of Soviet-rich. This implies that the average team size in Soviet-rich subfields did indeed increase disproportionately more relative to Soviet-poor subfields after the knowledge shock caused by the collapse of the Soviet Union, consistent with the theory.

The main concerns regarding causal interpretations of these results have to do with simultaneity, selection and omitted variable bias. The observed enlargement in authorship team size could be a result of a sudden liberalization in communication. We address this concern by removing journal articles published by Soviet mathematicians from our sample. Hence, the difference-in-differences estimation captures only increases in authorship team size directly attributable to coping with the surge in mathematical knowledge. We attempt to address selection concerns by including year and subfield fixed effects. To address concerns of omitted variable bias and preexisting time trends we estimate and plot coefficients of the interaction between year dummies and our measure of Soviet-rich. Specifically, we estimate:

$$TeamSize_{it} = \sum_{s=0}^{t} \beta_s \, (\gamma_s \times \, SovietRich_i) + \, Subfield_i + \gamma_t + \, \varepsilon_{it}$$

In Figure2 we plot the coefficients β_s along with confidence intervals for our main dataset where Soviet-rich is defined as top five ranking subfields and Soviet-poor is defined as bottom five ranking subfields of theoretical mathematics. $TeamSize_{it}$ is the logged count of authors for

each academic paper *i* published in year t. $SovietRich_i$ is an indicator variable equal to 1 if the publication *i* belongs to the treated group and 0 otherwise. We calculate the key explanatory variable as the interaction between year dummies γ_t and $SovietRich_i$: equal to 1 for those publications *i* that belong to the treated group of subfields and that were released in year t. We include subfield and time fixed effects, hence the individual term $SovietRich_i$ is not part of the estimating equation.

The coefficients β_s represent the estimated disproportionate increase in team size for Soviet-rich subfields over the 40 year period of interest. We observe that β_s values are close to zero before the collapse of the Soviet Union. However the β_s values enter an upward trend after the collapse of the Soviet Union in 1990.

We also estimate β_s values for Soviet-rich defined as top half ranking subfields of theoretical mathematics and Soviet-poor defined as bottom half ranking subfields of theoretical mathematics (Figure 3). We observe the same trend as with our main definition of Soviet-rich and Soviet-poor. Furthermore, these results remain robust to all other alternative definitions of Soviet-rich.

Overall, our results strongly suggest a rise in authorship team size directly attributable to an increase in knowledge burden.

6. (Table 2)

7. Discussion and Conclusion

Scholars are paying increasing consideration to evidence pointing to factors altering the organization of innovation activity with salient socioeconomic implications. Negative

implications for long-run growth (Jones, 2009) and increasing inequality (Jones, 2011) are predicated on a raise in knowledge burden, a factor which alters the nature of innovative activity towards an increased need for teamwork. While descriptive statistics are consistent with the knowledge burden hypothesis (e.g., Nobel laureates are older when they perform their great achievement, the age at first innovation increases, etc.; Jones, 2009; Jones, 2010), they are also consistent with other alternative explanations such as institutional changes (e.g., college requirements, journal publication processes; Stephan, 2012), reduced collaboration costs (Agrawal and Goldfarb, 2008), and increasing capital costs (Mowery et al, 2004). We make use of an instrument correlated with increased knowledge burden but not correlated with academic collaboration except indirectly through its effect on the knowledge frontier (the collapse of the Soviet Union) to provide more compelling evidence that a rise in knowledge burden leads to an increase need for teamwork. We find that academic collaboration increases faster in those subfields of theoretical mathematics where Soviet mathematicians made outstanding contributions before the fall of the Iron Curtain, consistent with the knowledge burden hypothesis.

References

(1) Agrawal, A., Goldfarb, A. 2008. Restructuring Research: Communication Costs and the Democratization of University Innovation. American Economic Review, Vol. 98(4), pp. 1578–1590.

(2) Borjas, G. J., Doran, K. B. 2012. The Collapse of the Soviet Union and the Productivity of American Mathematicians. NBER working paper.

(3) Dubois, P., Rochet, JC., Schlenker, M. 2011. Productivity and Mobility in Academic Research: Evidence from Mathematicians. Working paper.

(4) Ganguli, I. 2011. Saving Soviet Science: The Impact of Grants When Government R&D Funding Disappears. Working paper.

(5) Graham, L. R. 1993. Science in Russia and the Soviet Union: A short history. Cambridge University Press.

(6) Graham, L. R. , Dezhina, I. 2008. Science in the New Russia: Crisis, Aid, Reform. Indiana University Press.

(7) Grossman, J. W. 2002. The Evolution of the Mathematical Research Collaboration Graph. Congressus Numerantium, Vol. 158, pp. 201-212

(8) Jones, B. 2009. The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? Review of Economic Studies.

(9) Jones, B. 2010. As Science Evolves, How Can Science Policy? NBER Innovation Policy and the Economy, Vol. 11.

(10) Jones, B. 2011. The Knowledge Trap: Human Capital and Development, Reconsidered. NBER Working paper

(11) Mowery, D., Nelson, R., Sampat, B., Ziedonis, A. 2004. Ivory Tower and Industrial innovation, Stanford University Press.

(12) Newman, M. E. J. 2004. Coauthorship Networks and Patterns of Scientific Collaboration. Center for the Study of Complex Systems and Department of Physics, University of Michigan, Ann Arbor.

(13) Polyak, B.T. 2001. History of Mathematical Programming in the USSR: Analyzing the Phenomenon. Mathematical Programming.

(14) Tikhomirov, V. M. 2007. On Moscow Mathematics - Then and Now. Golden Years of Moscow Mathematics, Second edition, co-publication of the AMS and the London Mathematical Society.

(15) Sonnenwald, D. H. 2007. Scientific Collaboration. Annual Review of Information Science and Technology, Vol 41, Issue 1, pp. 643-681.

(16) Stephan, P. 2012. How Economics Shapes Science. Harvard University Press.

(17) Walsh, J. P., Bayma, T. 1996. Computer Networks and Scientific Work. Social Studies of Science Vol. 26, Issue 3, pp. 661-703.

(18) Wuchy, S., Jones, B., Uzzi, B. 2007. The Increasing Dominance of Teams in Production of Knowledge. Science 316, p. 1036.

Figure 1: Areas of Mathematics

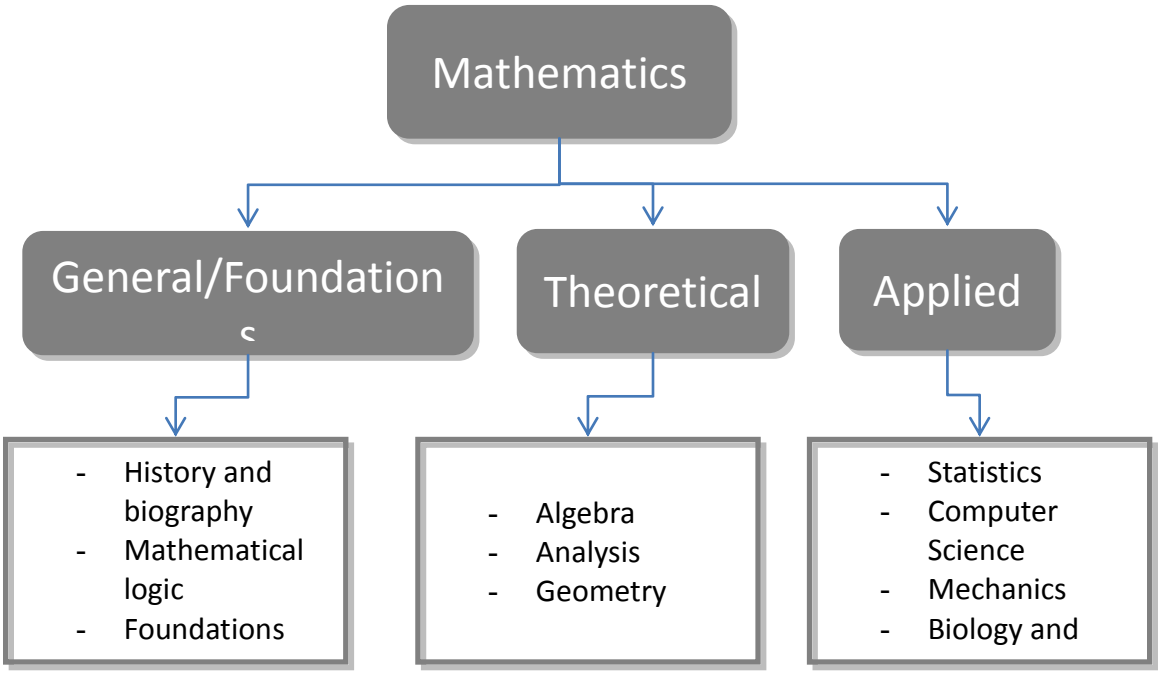


Figure 2: Disproportionate increase in mean team size for Soviet-rich subfields of mathematics, defined as top five ranking subfields of theoretical mathematics

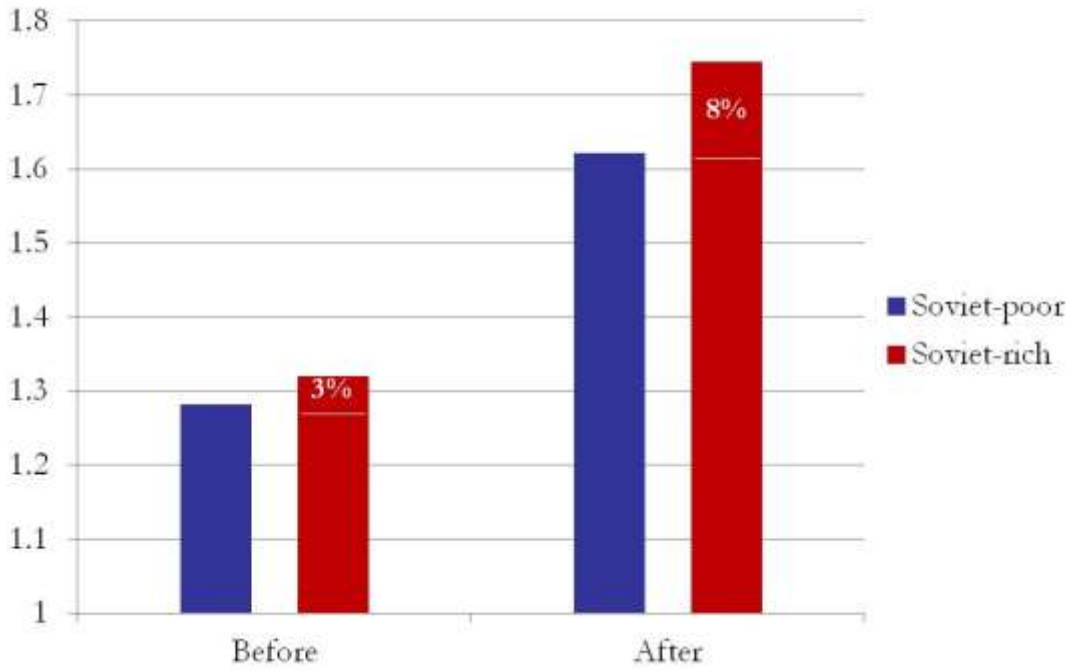


Figure 3: Disproportionate increase in team size for Soviet-rich subfields of mathematics, defined as top five ranking subfields of theoretical mathematics

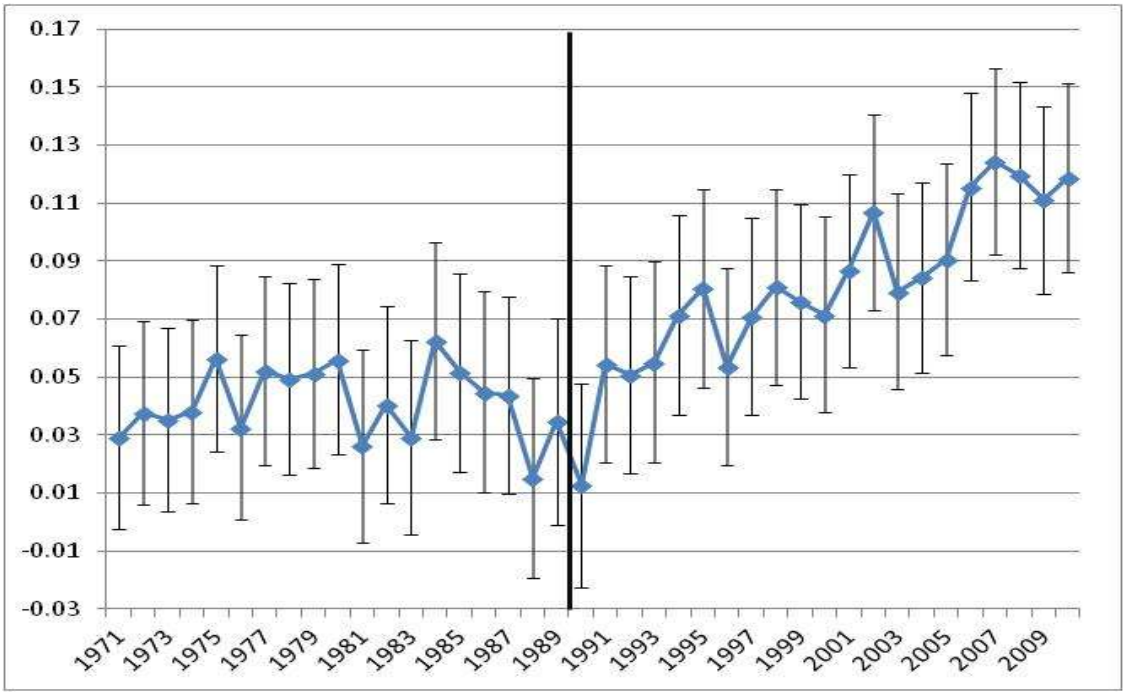


Figure 4: Disproportionate increase in team size for Soviet-rich subfields of mathematics, defined as top half ranking subfields of theoretical mathematics

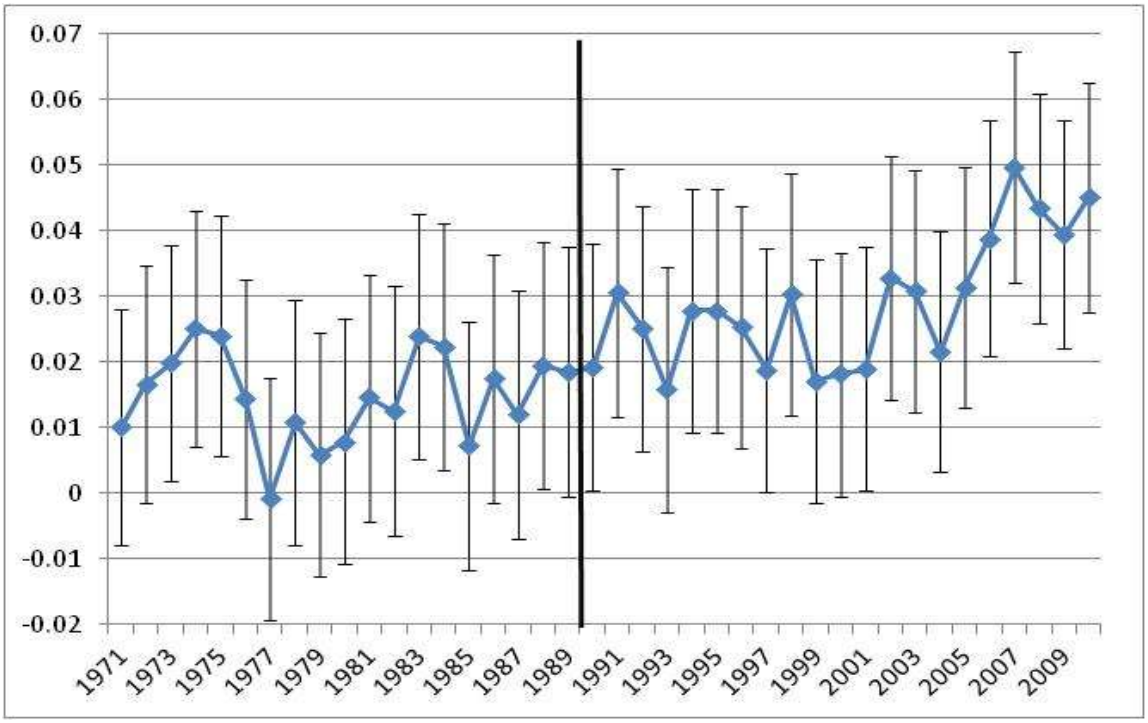


Table 1: Subfields of theoretical mathematics ranked based on proportion of published Soviet papers between 1970 and 1989

Ranking based on Soviet contribution (1970-1989)	MSC	Theoretical mathematics category	Description
1	35	Analysis	Partial differential equations
2	45	Analysis	Integral equations
3	47	Analysis	Operator theory
4	30	Analysis	Functions of a complex variable
5	42	Analysis	Fourier analysis
6	58	Geometry	Global analysis, analysis on manifolds
7	54	Geometry	General topology
8	39	Analysis	Difference equations and functional equations
9	49	Analysis	Calculus of variations and optimal control; optimization
10	22	Algebra	Topological groups, lie groups, and analysis upon them
11	17	Algebra	Non-associative rings and non-associative algebras
12	8	Algebra	General algebraic systems
13	32	Analysis	Several complex variables and analytic spaces
14	44	Analysis	Integral transforms, operational calculus
15	41	Analysis	Approximations and expansions
16	31	Analysis	Potential theory
17	52	Geometry	Convex geometry and discrete geometry
18	28	Analysis	Measure and integration
19	20	Algebra	Group theory and generalizations
20	33	Analysis	Special functions
21	18	Algebra	Category theory; homological algebra
22	26	Analysis	Real functions, including derivatives and integrals
23	5	Algebra	Combinatorics
24	55	Geometry	Algebraic topology
25	40	Analysis	Sequences, series, summability
26	51	Geometry	Geometry
27	14	Algebra	Algebraic geometry
28	6	Algebra	Order theory
29	12	Algebra	Field theory and polynomials
30	43	Analysis	Abstract harmonic analysis
31	13	Algebra	Commutative rings and algebras
32	57	Geometry	Manifolds
33	15	Algebra	Linear and multilinear algebra; matrix theory

Table 2: Difference-in-differences regression results

Difference-in-differences regression analysis					
Dependent variable: Log of author count per publication per year					
	Soviet-rich: top five	Soviet-rich: top three	Soviet-rich: top 25%	Soviet-rich: top 50%	Soviet-rich: reverse rank
AfterIronCurtain x SovietRich	**0.0507***** (0.0039)	**0.0512***** (0.0044)	**0.0289***** (0.0022)	**0.0159***** (0.0021)	**0.0014***** (0.0001)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes	Yes
R-squared	0.111	0.109	0.106	0.106	0.106
Observations	209,255	159,626	563,462	563,462	563,462
	All models are OLS with robust standard errors *significant at 10%, **significant at 5%, ***significant at 1%				